J. Vet. Med. B **53,** 1–10 (2006)

ISSN 0931–1793

Department of Biological Statistics and Computational Biology, College of Agriculture and Life Sciences, Cornell University, Ithaca, NY, USA

Human-mediated Foot-and-mouth Disease Epidemic Dispersal: Disease and Vector Clusters

A. L. Rivas[1,4], B. Kunsberg[2], G. Chowell[2], S. D. Smith[3], J. M. Hyman[2] and S. J. Schwager[1]

Addresses of authors: [1]Department of Biological Statistics and Computational Biology, College of Agriculture & Life Sciences, Cornell University, Ithaca, NY, USA; [2]Mathematical Modeling and Analysis, Los Alamos National Laboratory, Los Alamos, NM 87545, USA; [3]Institute for Resource Information Systems, College of Agriculture & Life Sciences, Cornell University, Ithaca, NY, USA; [4]Corresponding author: Tel.: +1 607 255 1644; fax: +1 607 255 4698; E-mail: alr4@cornell.edu

With 5 figures and 1 table

Received for publication August 5, 2005

Summary

Disease clusters were retrospectively explored at national level using a geo-referenced dataset from the 2001 Uruguayan Foot-and-Mouth Disease (FMD) epidemic. Disease location and time (first 11 epidemic weeks) were analysed across 250 counties (of which 160 were infected), without and with control for human mobility related factors (human population and road densities). The null hypothesis of random disease distribution over space and/or time was assessed with: (i) purely temporal; (ii) purely spatial; and (iii) space/time tests. At least within epidemic weeks 2 and 6, a principal disease cluster was observed in 33 contiguous counties ($P < 0.01$). Two secondary clusters, located at >100 km from each other, were also observed ($P < 0.01$). The purely spatial test that controlled for human population density identified two non-contiguous clusters ($P < 0.01$). Space and time analysis also revealed the same 33 counties as members of the principal cluster, of which 31 were also clustered when human population was controlled ($P < 0.01$). No clusters were reported by the spatial test when road density was assessed. The hypothesis that human mobility related factors autocorrelate with disease was empirically supported by two pieces of information: (i) removal of human population/road densities eliminated >93.9% of the counties included in the principal disease cluster; and (ii) statistically significant correlations ($P < 0.05$) were observed in the first three epidemic weeks between road density and the number of cases. Clusters where human population density was associated with 47% greater number of cases/sq. km than that of the principal cluster indicated possible roles as disease vectors (vector clusters). Selective control policy in vector clusters is recommended. Periodic (i.e. weekly) cluster and correlation analyses of both disease and other covariates may facilitate disease surveillance and help design space-specific control policy.

Introduction

Site-specific identification of geographical regions where disease prevalence is significantly greater than in neighbouring areas (disease clusters) is a major objective of epidemiological decision making (Ward and Carpenter, 2000). Such objective requires the use of geo-referenced data. In diseases of rapid dissemination (such as Foot-and-Mouth Disease or FMD), identification of disease clusters has been suggested to be of critical importance (Rivas et al., 2004; Chowell et al., 2006). However, although investigated at some locations (Wilesmith et al., 2003), FMD clustering has not yet been explored at national basis. While the role of road density in disease dispersal has been assessed before (Rivas et al., 2004), that study did not assess the national road network. Human demographics, although demonstrated to be associated with other diseases (Harrington et al., 2005), have not yet been investigated in relation to FMD epidemic dispersal.

Similar case prevalence may be the effect of close geographical proximity, which could also be expressed as similar case prevalence for locations 'close' in time (infected within a brief time interval). Those situations are mathematically expressed as spatial autocorrelation (as when case prevalence is similar for locations 'close' in space and/or time, Moran, 1950; Mantel, 1967). However, disease clusters may also involve dynamic interactions. Similar case prevalence may be the effect of connections that 'dissolve' both space and time: when communications are rapid and abundant, points far apart in space may be infected as well as (if not more than) points close to each other. Both human population and road densities, alone or combined, may interact and become a vector for disease dispersal (Harrington et al., 2005). Long distance spread may also occur as a result of multiple vulnerabilities, which may be trade related, and involve territorial areas far apart from each other, as observed in Great Britain in 2001 (Woolhouse and Donaldson, 2001).

While many tests may assess global disease spatial clustering (Alt and Vach, 1991; Besag & Newell, 1991; Cuzick and Edwards, 1990; Diggle & Chetwynd, 1991; Grimson, 1991; Moran, 1950; Ranta, 1996; Tango, 2000), they cannot indicate the specific location of disease clusters; in contrast, Kulldorff's SaTScan test does (Abrial et al., 2003; Doherr et al., 2002; Sheridan et al., 2005). SaTScan provides a 'cylinder' of varying diameters that scans a surface. Its circular base measures space, while its height represents time (Kulldorff et al., 2005).

The FMD epidemic that affected Uruguay in 2001 provides an opportunity to explore disease clustering at national level. In this epidemic: (i) its index case affected bovines (a species that usually presents with clinical signs); and (ii) its index case was reported at a farm level, which indicates the epidemic was noticed early. At least these two factors differed from those observed in the 2001 British FMD epidemic, where the

predominant species infected by FMD virus was sheep (usually, an asymptomatic host), and the first case was reported at slaughterhouse level (i.e. much later in the epidemic progression) (Woolhouse and Donaldson, 2001).

Using geo-referenced and temporal data from the 2001 Uruguayan FMD epidemic, this study was set to explore: (i) whether disease clusters may be observed in rapidly disseminating diseases; and (ii) to assess the role of human demographics and/or road density as covariates of disease clustering and spread.

Materials and Methods

Data sources

A 1 : 500 000 scale political division geographical chart of Uruguay, kindly provided by its producer (the Geographical Service of the Uruguayan Ministry of Defense, http://www.ejercito.mil.uy/cal/sgm/frame3.htm), was used as the basic map. It was geo-referenced into GIS software (ESRI, Redland, CA, USA), providing national, state and county border contours as well as the national highway network. Data on human population were retrieved from the 1996 national census (http://www.ine.gub.uy/). Although the human population of interest was that of the first half of 2001, the mean annual growth of the Uruguayan population (between 1996 and 2004) has been 0.35 of a percent point (http://www.ine.gub.uy/). Therefore, the error associated with this data source was estimated to be < 0.02. Data on case location and time ($n = 1721$ farms, located in 160 counties and infected over 11 epidemic weeks), were aggregated at national level ($n = 250$ counties, where 37 818 farms were assumed to be active at the time of this epidemic). Data on FMD infected farms (cases) were obtained from public records of the Uruguayan Ministry of Livestock, Agriculture and Fisheries (MGAP, http://www.mgap.gub.uy) between 20 June and 17 July 2002. The *2000 Annals* and *2000 Agricultural Census* provided additional spatial data on farms (http://207.3.127.35/Diea/anuarios.htm, http://207.3.127.35/Diea/default.htm). Information on this epidemic has been provided elsewhere [Reports 3342 and 3456 (2001) of the Food and Veterinary Office of the European Commission's Health and Consumer Protection-Directorate General, http://europa.eu.int/comm/food/fs/inspections/vi/reports/uruguay (reports 3342-2001 and 3456-2001), Rivas et al., 2003a,b, 2004].

Methods

Several spatial layers were built. The foundational layer contained three variables: (i) county case prevalence (on weekly basis, for 11 weeks beginning on 23 April 2001); (ii) human population density; and (iii) road network. The area (sq. km) of individual and clustered counties was directly calculated by GIS. Two secondary variables were also created by GIS: (i) county human population (by adding the population of villages and cities within a county and assigning that value to the corresponding county surface); and (ii) road density (by intersecting highway data with county surfaces).

Software and procedures

Digital and graphical data were geo-referenced and processed using ArcGIS 8.x and ArcView 3.x (both from ESRI, Redland, CA, USA). *Queries* were performed and new sets were created using the *add a new set* command, which included counties identified as clustered. The *new sets* so identified by queries were then converted into new datasets. Disease clustering was assessed with a spatial statistics package for cluster identification (SaTScan^{TM}, version 5.1.1, http://www.satscan.org/).* Analysis was complemented with a statistical package (Minitab 14, Minitab Inc., State College, PA, USA).

Statistical analysis

The number of cases per county, their spatial location and the time they occurred were assumed to be Poisson distributed. The null hypothesis was that the number of infected farms per county, their location and time was similar across counties and time. Therefore, the disease likelihood ratio was expected to be constant over the whole territory and time frame under analysis. Analysis consisted of three steps. (i) The hypothesis that the likelihood ratio for the number of cases reported within a certain geographical area was higher than that observed outside that area at some time during the epidemic was assessed by a purely spatial test. A purely spatial test investigated whether case clustering occurred at all throughout the epidemic. (ii) Clustering throughout different time intervals was assessed by the space–time test. (iii) And the role of covariates (county human population and county road densities) was investigated in spatial, and in space and time analyses.

These tests were conducted using SaTScan (Kulldorff and Nagarwalla, 1995; Kulldorff, 1997) which imposes a circular window that defines zones over the territory under analysis. When the window contains a given county's centroid, that county is included in the analysis (Kulldorff et al., 1997). The circular window's centre is then moved over the whole space so that different sets of neighbouring counties are investigated. The centre of the window was positioned at the centroids of all 250 counties in Uruguay. The radius of the circular window varied continuously from zero to a maximum radius so that the window's coverage would not exceed 50% of the total farm population. The most likely cluster was determined by maximizing a likelihood function over all the zones. The significance of the most likely cluster was evaluated through Monte Carlo simulation (9999 repetitions) (Kulldorff and Nagarwalla, 1995; Kulldorff, 1997). A likelihood ratio was created by dividing the maximum likelihood value by another likelihood value based on the null hypothesis. For example, if the likelihood ratio of the potential cluster ranked in the top 500 likelihood ratios out of 9999 simulations, we would say that the most likely cluster was significant ($P < 0.05$). Secondary clusters could be calculated similarly.

The correlation coefficient between the rates of epidemic growth in each infected county and covariates (road and human population densities) was calculated at different time periods. The growth rates m_k ($k = 1$ to 163 counties) for each county were defined by the slope of the line that best fitted the cumulative number of weekly cases at a specified time period.

*SaTScan^{TM} is a trademark of Martin Kulldorff. The SaTScan^{TM} software was developed under the joint auspices of Martin Kulldorff of the National Cancer Institute and of Farzad Mostashari at the New York City Department of Health and Mental Hygiene.

We considered epidemic periods of length 2, 3, 4 and 5 weeks starting from week 1. For example, the epidemic periods of length 2 over the eleven epidemic weeks would be (1,2), (2,3), (3,4), (4,5), (5,6), (6,7), (7,8), (8,9), (9,10) and (10,11). Each of these epidemic periods had an associated growth rate for each county. The resulting growth rates in each county were then correlated with the county road or human population density. For a specific time period length (e.g. 2 weeks), a sequence of correlation coefficients (and their corresponding *P*-values) was obtained for each covariate.

Results

The epidemic began in the southwestern region of Uruguay. A discontinuous cluster-like epidemic structure was suggested by the data since its first epidemic week (Fig. 1). Significant correlations were observed between road density and human population density, and (in the first 3 epidemic weeks) between road density and the cumulative number of county cases (expressed as case growth rate, $P < 0.05$, Figs 2 and 3). After epidemic week 3, road density ceased to be a predictor of epidemic cases. Human population density, although not reaching statistical significance, also showed similar patterns.

A purely temporal test rejected the null hypothesis, indicating the presence of disease clustering in the first six epidemic weeks ($P < 0.01$, not shown). The purely spatial test revealed three non-contiguous disease clusters. The principal cluster (composed of 33 counties) followed the southwestern coast of the country (Fig. 4a). Two additional clusters were also observed.

The space and time test identified the same 33 counties as the principal cluster and the same 6 counties previously identified in cluster 'B' by the spatial test. In addition, the space and time test detected more counties in cluster 'C' than the spatial test did (Fig. 4a and b). This resulted from the fact that in the space–time test, smaller time intervals were investigated and, therefore, clustered locations had relatively higher case concentrations, which yielded a significantly greater likelihood ratio (Table 1). The principal cluster was statistically significant between epidemic week 2 and the end of week 6. The

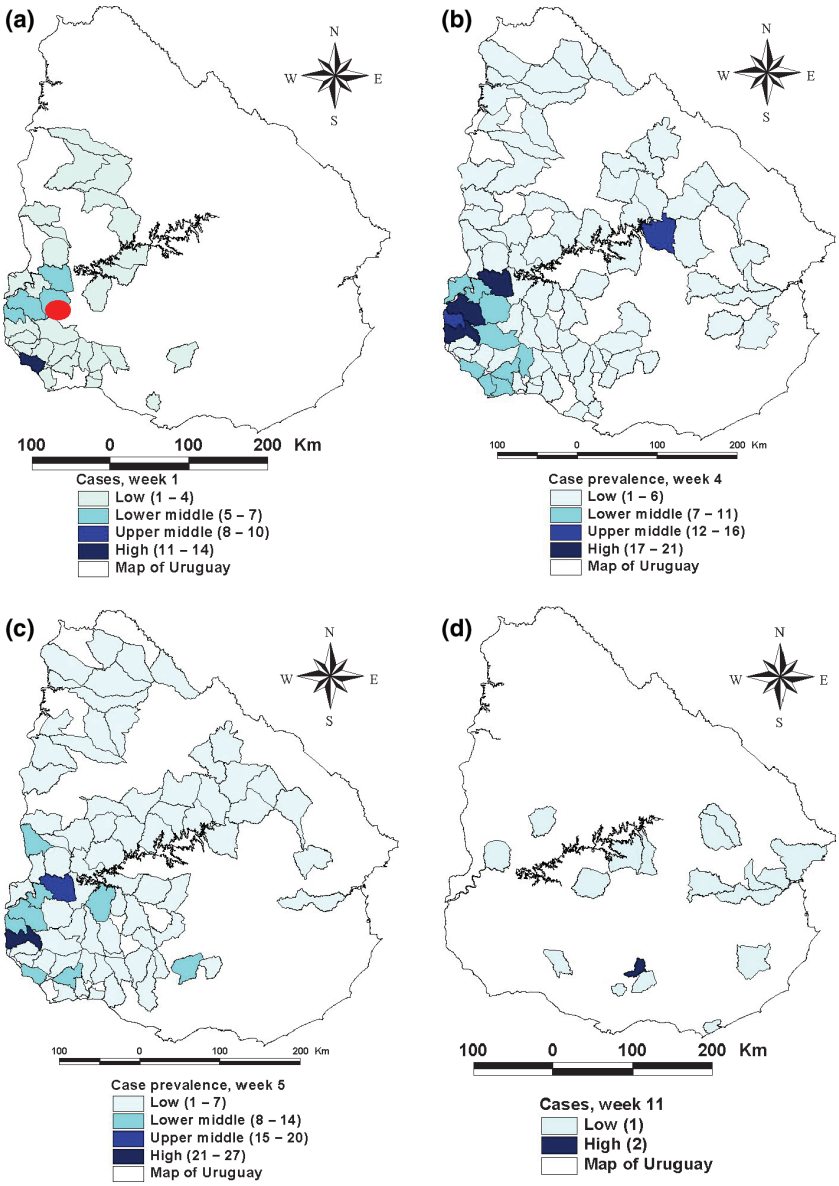
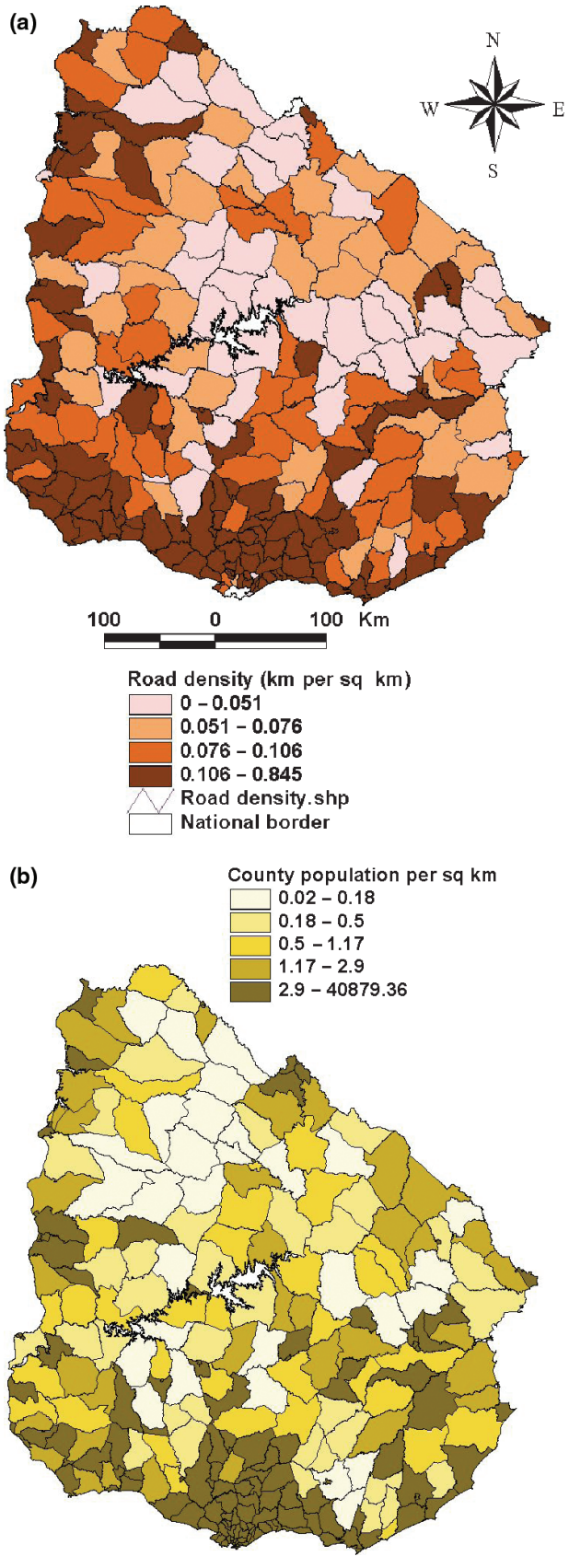


Fig. 1. Map indicating the location of Uruguay where on 23 April 2001, a Foot-and-Mouth (FMD) epidemic was reported. The point where the index case was observed is shown with a red circle. The county disease prevalence (number of infected farms) is expressed in three levels of equal weight (low, middle and high), proportional to the cases reported on each week. (a–d): epidemic weeks 1, 4, 5 and 11.

largest secondary cluster (cluster 'B') occurred between weeks 4 and 8. Cluster 'C' was significant between weeks 4 and 7 (Fig. 4b). The three clusters were significant at $P = 0.01$.



When human population was assessed by a purely spatial test, 31 of the original counties were excluded. Only two counties remained in the principal disease cluster (Fig. 4c).

A single disease cluster was identified by the space and time test that controlled for human population density. It included 31 of the 33 counties reported by the space and time test that did not control for covariates (Fig. 4d).

The spatial test that controlled for road density covariation did not indicate clustering (either alone or combined with human population, Table 1). In contrast, the space and time test that controlled for road density revealed an extensive cluster that comprised most of the western border (Fig. 5a). Identical results were obtained when both road and human population densities were controlled (Fig. 5b). However, the farm prevalence revealed by this extensive cluster was the lowest (0.0118 cases/sq. km), a value that represented only 26% of that of the principal cluster (0.0456 cases/sq. km, Table 1). In spite of its extension, the 70-county cluster did not include 12 of the 33 counties identified in the principal cluster (Fig. 5b). Correlation analyses revealed that in the early weeks road density correlated significantly with the number of cases (epidemic growth rate, Fig. 3b).

Simultaneous analysis of all spatial tests indicated that one of the two clusters that controlled for human population density (including counties #1703 and 1707) was located inside the principal disease cluster. The other cluster where human population was controlled was located outside the principal disease cluster (Fig. 5c). The cluster including counties #1703 and #1707 displayed 0.0673 cases/sq. km, which represented a 47% greater number of cases/sq. km than that shown by the principal disease cluster (0.0456, Table 1).

Discussion

Summary and interpretation

While based on geo-referenced and temporal data of an actual epidemic, this report should not be construed as an evaluation of the FMD epidemic that occurred in Uruguay in 2001, but as a hypothetical (although realistic) scenario that facilitates the retrospective exploration and/or generation of hypotheses on epidemic spread. The findings and generated hypotheses are limited to the dataset investigated here. Because it is based on data that do not necessarily correspond to the time frame of interest (the April to July 2001 human population density) and lacks data that may have been critical (actual human traffic), findings should not be regarded as evidence of causation but, at best, associations that may support hypotheses.

The data were compatible with two major hypotheses: (i) disease clusters may be observed even in rapidly disseminating epidemics, such as those caused by FMD virus; and (ii) human mobility related factors (demographic and road densities) may facilitate epidemic dispersal (vector clusters).

Evidence of disease clustering was continuously observed at least until the end of the sixth epidemic week. Although the rapid dissemination of FMD might result in spatial changes

Fig. 2. County road density (a) and human population density (b). County road density is expressed as the (square) county interstate road length/county perimeter (km/km). Human population is expressed as the (log) number of county inhabitants/county area (sq. km), as reported in the 1996 census.

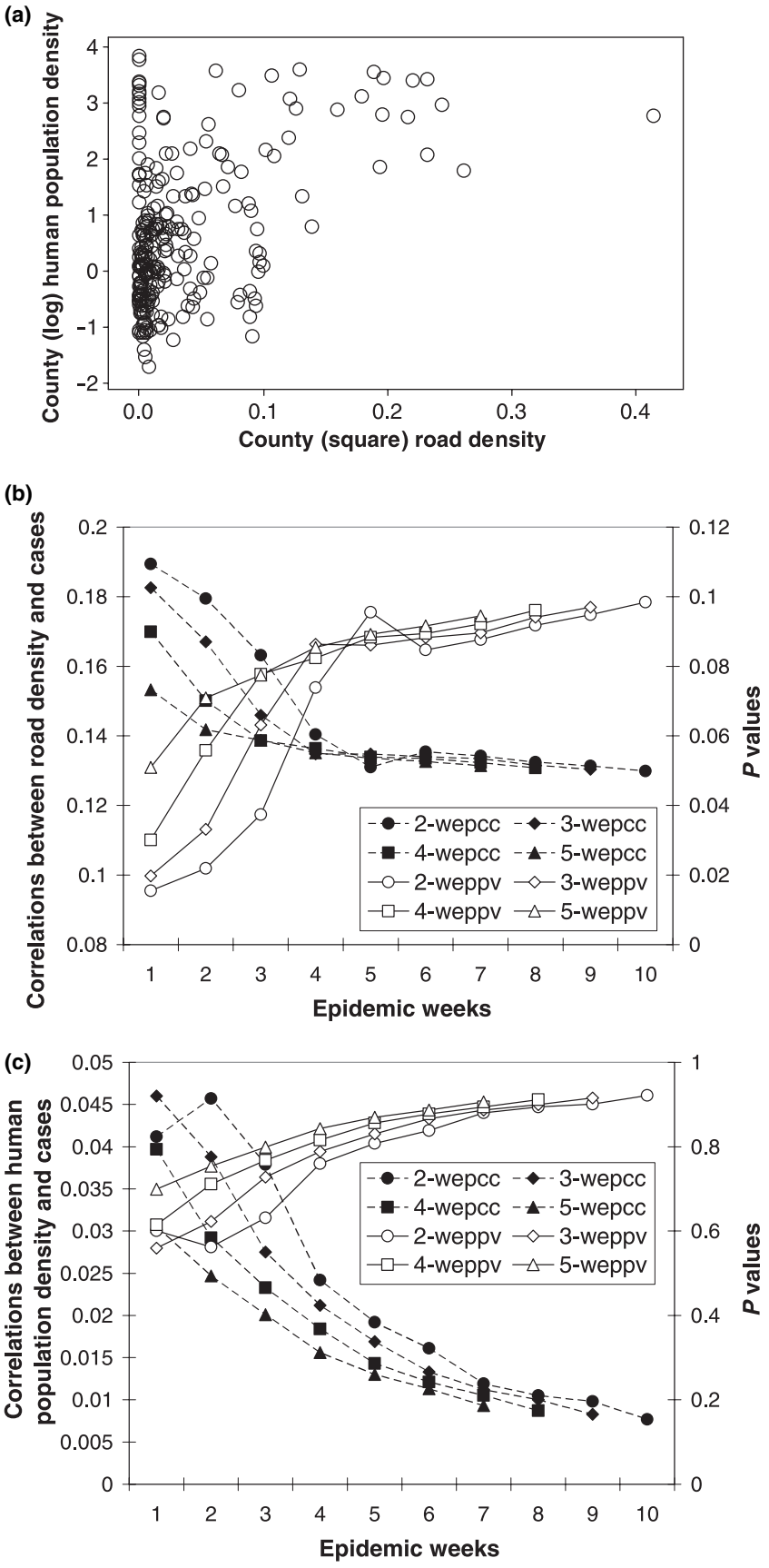


Fig. 3. Correlations between county road density and county human population density, and between covariates and the number of county cases (epidemic growth rate). The correlation between county road density and county human population density is displayed in (a), that of road density is shown in (b) and human population density is shown in (c). The correlation coefficient between road and human population densities was $r = 0.43$ [$P < 0.001$, $n = 249$ counties, (a)]. Highly urbanized counties (i.e. those of the state of Montevideo) were not included. The correlation coefficient between covariates and growth rate at specified epidemic time intervals [week epidemic period (i.e. 2-, 3-, 4- or 5-week epidemic period), is shown in closed symbols as [2–5] week epidemic period correlation coefficient or *wepcc*], whereas the corresponding *P*-values are shown in open symbols [i.e. (2–5) week epidemic period *P*-value or *weppv*] (b, c). See text for the description of epidemic growth rate.

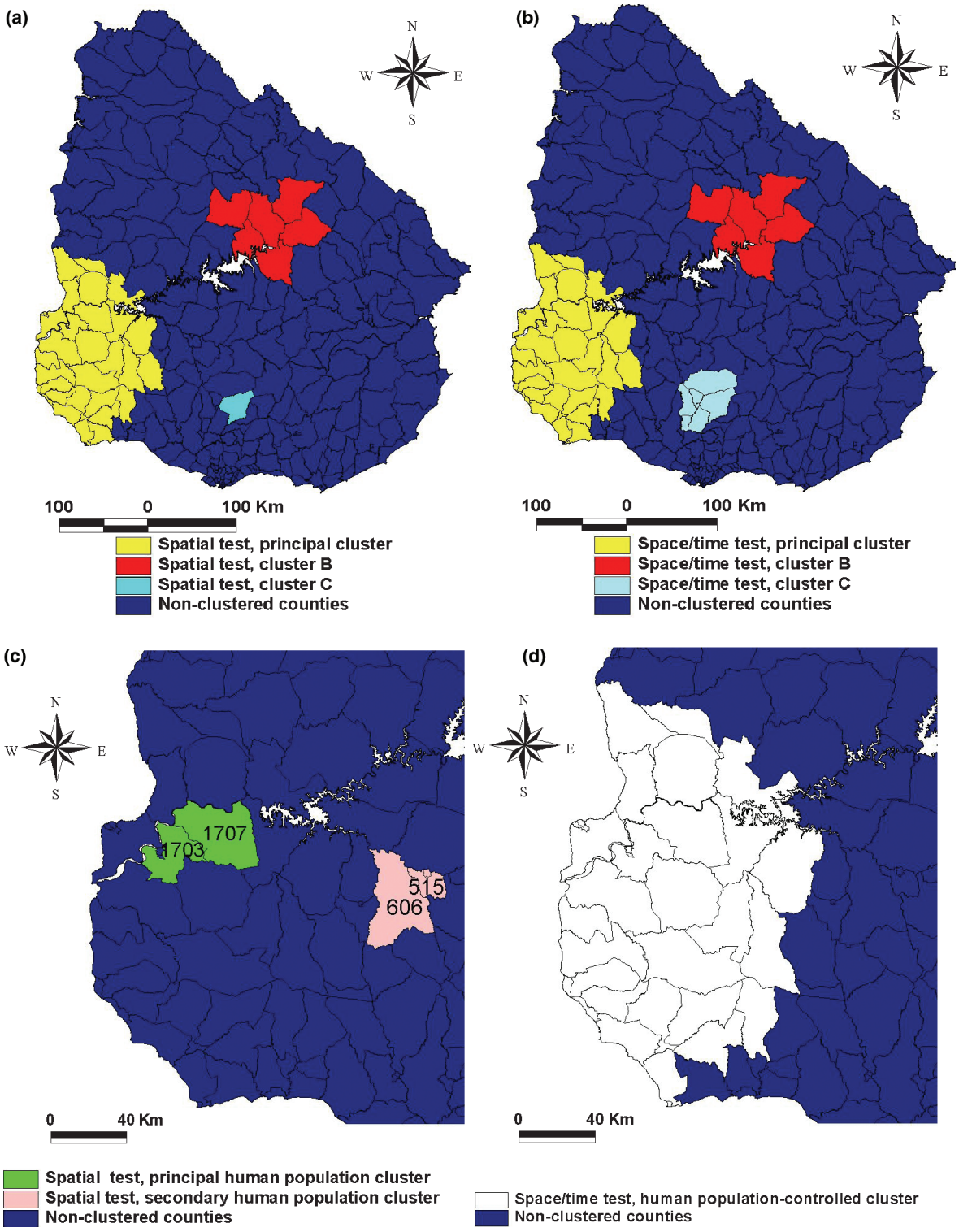


Fig. 4. Spatial and space/time analysis of disease clusters without and with control for human population density. Without controlling for covariates, a principal and two secondary disease clusters are indicated by the spatial test [$P < 0.01$, (a)]. A similar, although partially larger disease clustering is reported by the spatial and time analysis (b). After controlling for county human population density, two clusters where the number of cases is greater than in counties of similar population density are indicated by the spatial test (c). The space and time test that controlled for human population density included all but two of the counties reported as clustered by the non-controlled spatial/temporal test (d).

over time, the data suggested that disease clusters were structured soon after the epidemic onset took place and remained so for over a month. Such finding may be particular of the epidemic under analysis (not generalizable to all epidemics). However, in situations where the index case is reported at a farm level and the species affected is symptomatic (as occurred in the scenario under study), cluster analysis may have the potential to become a tool applicable to design and/or

Table 1. Disease clusters as determined by the purely spatial and the space and time tests

Test	Counties in cluster	Area (sq. km) covered by (major/minor) cluster	Cases in cluster	Cases per sq. km	Ratio above or below mean[a]	Cases expected in cluster	Cluster likelihood ratio	Cut-off for $P = 0.01$	Number of non-farm counties[b]
Temporal	160	137 820.3	1278	0.0093	1.00	772.9	305.6	5.98	20
Spatial	33	20 062.3 (major)	916	0.0456	3.65	188.4	930.4	9.9	4
Spatial	6	8788.8 (minor)	115	0.0130	1.04	40.8	46.7	9.9	0
Spatial	1	849.2 (minor)	30	0.0353	2.82	6.9	21.9	9.9	0
Space/time	33	20 062.3 (major)	758	0.0377	4.09	84.6	1151.6	13.3	4
Space/time	6	8788.8 (minor)	94	0.0107	1.35	18.7	78.3	13.3	0
Space/time	10	6209.7 (minor)	84	0.0135	1.84	28.3	36.6	13.3	1
Spatial, HPD controlled	2	1619.5 (major)	109	0.0673	5.38	63.0	14.4	9.4	0
Spatial, HPD controlled	3	1287.7 (minor)	11	0.0085	0.68	1.6	11.7	9.4	2
Space/time, HPD controlled	31	18 676.0	730	0.0391	4.25	382.9	173.4	12.8	4
Spatial, RD[d] controlled	0	–	–	–		–	–	–	–
Space/time, RD controlled	70	61 046.3	725	0.0118	1.28	384.4	167.0	12.8	12
Spatial, HPD and RD controlled	0	–	–	–		–	–	–	–
Space/time, HPD and RD controlled	70	61 046.3	725	0.0118	1.28	384.4	167.0	12.8	12

HPD, human population density; RD, road density.
[a]The total area covered by the epidemic in the first 11 weeks was 137 820.26 sq. km (160 counties), where 1721 farms were reported to be infected (national mean 0.0125 cases/sq. km). The ratio above/below the national mean is adjusted to the total number of cases reported in the country at the timeframe each test was significant (see text).
[b]Counties with no cases (occupied by cities) or false positive disease cluster results.

evaluate control policy. For example, repeated (i.e. weekly) cluster analyses could had been used to weigh the costs and benefits of different policies (i.e. stamping-out versus national post-outbreak vaccination). The data suggested different microregional susceptibilities. For instance, counties in the centre of the country were not included in any disease cluster at any time (Fig. 4a and b). In spite of being surrounded by disease clusters on three sides, the very centre of the country remained marginally affected by the epidemic. A possible explanation is the fact that the centre of the country is occupied by a major river and a lake, which may act as natural barriers of disease spread (Smith et al., 2002).

This study confirmed, with national-level data, a previous report indicating that road density may be associated with disease dispersal (Rivas et al., 2004). An autocorrelation between road density and disease clustering was suggested by the fact that none of the 33 counties identified as clustered by the spatial test was so identified when road density was controlled. Human population density also provided an indication of autocorrelation with disease: two of the counties identified to belong to the principal disease cluster were eliminated when demographic density was controlled. Although the hypothesis that human mobility related factors may facilitate epidemic spread was supported empirically (Fig. 3b), lack of data on actual human traffic during this epidemic prevented confirmation of such hypothesis.

Different levels of influence on epidemic spread were also provided by cluster analysis. For instance, the influence of road density seemed to be greater than that of human population density (i.e. no disease cluster was observed when road density was controlled for). In addition, statistically significant correlations were observed between road density and epidemic growth rate in the first 3 weeks.

The hypothesis that human mobility related factors may facilitate epidemic dispersal was supported by the results observed when possible confounders were controlled (i.e. by the amount of disease clustering removed when covariates were assessed). Conversely, information potentially applicable by control policy was produced by results observed after the influence of covariates was removed (the amount of disease remaining after accounting for covariates).

However, before applications in control policy are explored, it is necessary to recall that the type of disease being studied differs markedly from cancer or chronic infections (i.e. *Bacillus anthracis*, paratuberculosis), diseases assessed in the past with this analytical approach (Kulldorff et al., 1997; Smith et al., 2000; Ward and Pérez, 2004). Unlike those, FMD spreads very rapidly which results in a feature not observed in slowly spreading diseases: factors associated with FMD clustering may be either inside or outside where disease is observed.

The previous concept can be graphically conveyed by noticing that 'neither cows are seen in major cities nor skyscrapers are observed within farms'. And yet, they may be connected within a short time interval. People moving across highways that connect human enclaves (and go across counties of lower human population density, where farms are located) may be acting as disease vectors even if they do not reside in farm areas. As a result, if analyses only focus on where disease is observed, researchers could miss what is 'outside the picture' (although may be causing that picture). Vector clusters may be inside or outside the space where disease clustering occurs. It is suggested that cluster analysis of rapidly disseminating diseases should not be used in the same way it is used when time is ignored or not regarded as a biologically relevant factor.

However, the previous hypothesis should not be construed as synonymous of *sufficient* causation. Conceptually, causes may be classified as *necessary* and *sufficient*. Although a specified outcome (i.e. infection) cannot occur in the absence of a necessary cause, the mere presence of such cause does not inexorably lead to the specified outcome. It may also occur that the final outcome is the result of multiple factors, of which some may act as facilitators and some as obstacles (Smith et al., 2002).

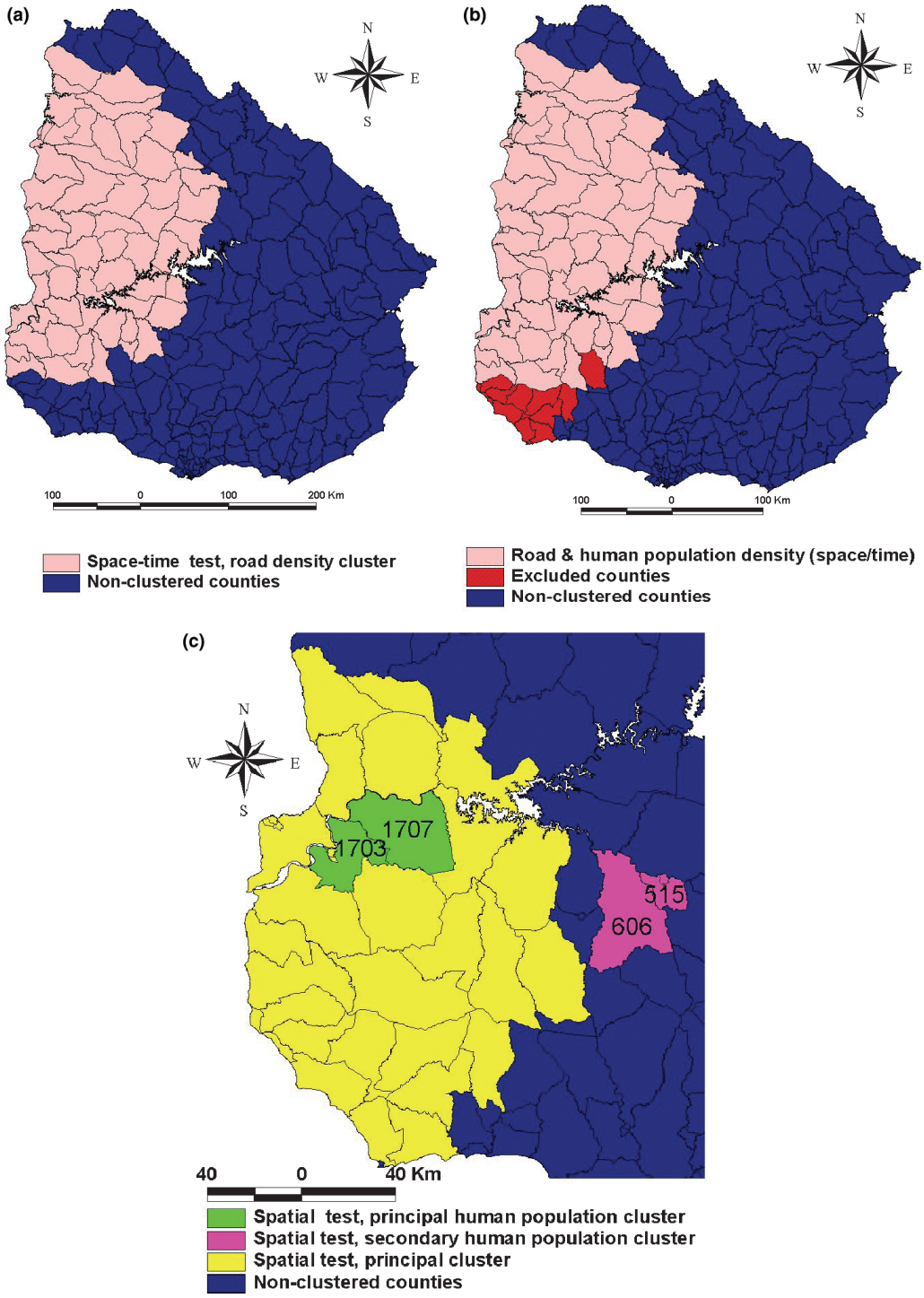


Fig. 5. Possible applications for control policy. A space and time test that controlled for road density shows a 70-county cluster (a). An identical, 70-county cluster is reported by the spatial/temporal test when both road and human population densities are controlled, however, 12 counties (included in the principal disease cluster) are excluded (b). The 33-county principal disease cluster so reported by both the spatial and the space/time tests is shown in relation to the two human population density-controlled clusters [one within, the other outside the principal disease cluster, (c)]. Selective control policy of vector clusters [i.e. counties 1703 and 1707, (c)] would result in at least 47% greater number of cases being prevented than controlling for the principal disease cluster (see Table 1).

Therefore, it is suggested that cluster analysis of rapidly disseminating diseases may be used with emphasis not only on where disease is observed (disease clusters), but also on where putative associated factors are located (even if they are outside the disease cluster), without assuming that such factors (vector clusters) represent sufficient causes of epidemic spread. Control policy may weigh the potential costs and benefits of adopting measures that could be effective only if those factors were sufficient causes (which, usually, are unknown).

The previous considerations apply to the situation observed in counties identified as #1707 and 1703; the cluster located outside the principal disease cluster (Fig. 5c); and the large, 70-

county cluster reported by the tests controlling for road density, which did not consider 12 counties included in the principal cluster (Fig. 5b). When human population density was controlled, counties #1707 and 1703 revealed greater disease prevalence than other counties of similar human population density and even a 47% greater number of cases/sq. km than that observed in the principal disease cluster (0.067/0.0456, Table 1 and Fig. 5c). Because this cluster was located inside the region where the principal disease cluster was reported, it is plausible to postulate that selective control policy of counties #1707 and 1703 (i.e. thorough vehicle disinfection on all outgoing traffic) might have diminished regional epidemic spread. Those two counties appeared to act both as disease and vector clusters.

A second cluster was located outside (although close to) the principal disease cluster, which included counties #606, 515 and 501 (Fig. 5c). It revealed a much lower number of cases/sq. km (even lower than national average, Table 1), which might disqualify that cluster for selective control policy. However, considering that such information was not available to policy makers during the early phases of the epidemic, and assuming that the cost of selective control policy in this cluster was marginal and could be implemented promptly (i.e. vehicle disinfection or traffic ban), it may be acceptable to err on the side of inclusiveness and regard this cluster as a possible vector cluster. Selective control policy (i.e. control of specific cross-roads associated with disease clusters) may be recommended if it is rapidly implementable and supported by a cost–benefit analysis, even when there is no evidence that control in such areas will prevent epidemic spread.

In contrast, it is suggested that the cluster reported by the space and time tests that controlled for road density did not justify a selective policy (Fig. 5b). It was the largest cluster (three times larger than the principal disease cluster, Table 1) and yet, it did not include 12 counties already determined to be part of the principal disease cluster.

Because cluster analysis is likely to overestimate disease clustering (Pérez et al., 2002) and rapidly disseminating diseases (such as FMD) do not necessarily share the biological dynamics observed in other diseases, it is suggested that use of spatial analysis requires simultaneous consideration of biological as well as mathematical factors. The combined use of disease and vector cluster analysis is proposed.

Problems and further research

The technique used here is based on the construction of a 'cylinder'. That is, its base is a circle. As a result, this technique is prone to merge discontinuities occurring within that circle (Tango and Takahashi, 2005), and ignore relevant factors within the reported cluster that may facilitate (or prevent) epidemic spread. For instance, this technique does not discriminate between urban and non-urban (farm) areas. Consequently, a reported disease cluster may contain urban counties (composed of cities, not farms). For example, four exclusively urban (non-farm) counties were observed within the principal cluster reported here and 12 such counties were included in the 70-county cluster generated when road density was controlled. Inclusion of such counties, due to the algorithm on which this approach is based, may lead to false positive results.

The validity of these findings could have increased if more variables and/or smaller scales had been considered. Assessment of case prevalence at farm level (as opposed to data aggregated at county level) might have been more informative (Pfeiffer, 1996).

Further integration between GIS and spatial statistics is suggested. The pursuit of testing approaches that attempt to identify disease clusters based on their actual shapes (not necessarily circular but, more probably, shaped as irregular polygons) is recommended. Periodic assessment of correlations between the epidemic growth rate and road density, of time length similar to the estimated reproduction cycle of the infective agent (in the case of FMD, every 3 days), if validated, may provide an additional policy (i.e. identification of specific cross-roads that may influence epidemic spread) to control FMD epidemics, which complements those reported before (Rweyemamu and Astudillo, 2002).

References

Abrial, D., D. Calavas, N. Lauvergne, E. Morignat, and C. Ducrot, 2003: Descriptive spatial analysis of BSE in western France. Vet. Res. **34,** 749–760.

Alt, K. W., and W. Vach, 1991: The reconstruction of 'genetic kinship' in prehistoric burial complexes-problems and statistics. In: Bock, H. H., and P. Ihm (eds), Classification, Data Analysis, and Knowledge Organization. Springer Verlag, Berlin.

Besag, J., and J. Newell, 1991: The detection of clusters in rare diseases. J. R. Statist. Soc. A **154,** 143–155.

Chowell, G., A. L. Rivas, S. D. Smith, J. M. Hyman, 2006: Identification of case clusters and counties with high infective connectivity in the 2001 Uruguayan foot-and-mouth disease epidemic. Am. J. Vet. Res. **67,** 102–113.

Cuzick, J., and R. Edwards, 1990: Spatial clustering for inhomogeneous populations. J. R. Statist. Soc. B **52,** 73–104.

Diggle, P. J., and A. G. Chetwynd, 1991: Second-order analysis of spatial clustering for inhomogeneous populations. Biometrics **47,** 1155–1163.

Doherr, M. G., A. R. Hett, J. Rufenacht, A. Zurbriggen, and D. Heim, 2002: Geographical clustering of cases of bovine spongiform encephalopathy (BSE) born in Switzerland after the feed ban. Vet. Rec. **151,** 467–472.

Grimson, R. C., and R. D. Rose, 1991: A versatile test for clustering and a proximity analysis of neurons. Meth. Inform. Med. **30,** 299–303.

Harrington, L. C., T. W. Scott, K. Lerdthusnee, R. C. Coleman, A. Costero, G. G. Clark, J. J. Jones, S. Kitthawee, P. Kittayapong, R. Sithiprasasna, and J. D. Edman, 2005: Dispersal of the dengue vector *Aedes aegypti* within and between rural communities. Am. J. Trop. Med. Hyg. **72,** 209–220.

Kulldorff, M., 1997: A spatial scan statistic. Commun. Statist. Theory M. **26,** 1481–1496.

Kulldorff, M., and N. Nagarwalla, 1995: Spatial disease clusters: detection and inference. Stat. Med. **14,** 799–810.

Kulldorff, M., E. J. Feuer, B. A. Miller, and L. S. Freeman, 1997: Breast cancer clusters in the Northeast United States: a geographic analysis. Am. J. Epidemiol. **146,** 161–170.

Kulldorff, M., R. Heffernan, J. Hartman, R. M. Assunção, and F. Mostashari, 2005: A space–time permutation scan statistic for the early detection of disease outbreaks. PloS Med. **2,** 216–224.

Mantel, N., 1967: The detection of disease clustering and a generalized regression approach. Cancer Res. **27,** 209–220.

Moran, P. A. P., 1950: Notes on continuous stochastic phenomena. Biometrika **37,** 17–23.

Pérez, A. M., M. P. Ward, P. Torres, and V. Ritacco, 2002: Use of spatial statistics and monitoring data to identify clustering of bovine tuberculosis in Argentina. Prev. Vet. Med. **56,** 63–74.

Pfeiffer, D. U., 1996: Issues related to handling of spatial data. In: McKenzie, J. (ed.), Proceedings of the Epidemiology and State Veterinary Programmes. New Zealand Veterinary Association/ Australian Veterinary Association, Second Pan Pacific Veterinary Conference, Christchurch, New Zealand 23–28 June, pp. 83–105. Christchurch, New Zealand.

Ranta, J., J. Pitkäniemi, M. Karvonen, E. Virtala, J. Rusanen, A. Colpaert, A. Naukkarinen, and J. Tuomilehto, 1996: Detection of overall space-time clustering in a non-uniformly distributed population. Stat. Med. **15,** 2561–2572.

Rivas, A. L., S. E. Tennebaum, J. P. Aparicio, A. L. Hoogesteijn, H. O. Mohammed, C. Castillo-Chávez, and S. J. Schwager, 2003a: Critical response time (time available to implement effective measures for epidemic control): model building and evaluation. Can. J. Vet. Res. **67,** 307–311.

Rivas, A. L., S. D. Smith, P. J. Sullivan, B. Gardner, J. P. Aparicio, A. L. Hoogesteijn, and C. Castillo-Chávez, 2003b: Identification of geographic factors associated with early spread of foot-and-mouth disease. Am. J. Vet. Res. **64,** 1519–1527.

Rivas, A. L., S. J. Schwager, S. Smith, and A. Magri, 2004: Early and cost-effective identification of high risk/priority control areas in foot-and mouth disease epidemics. J. Vet. Med. B **51,** 263–271.

Rweyemamu, M. M., and V. M. Astudillo, 2002: Global perspective for foot and mouth disease control. Rev. Sci. Tech. **21,** 765–773.

Sheridan H. A., G. McGrath, P. White, R. Fallon, M. M. Shoukri, and S. W. Martin, 2005: A temporal-spatial analysis of bovine spongiform encephalopathy in Irish cattle herds, from 1996 to 2000. Can. J. Vet. Res. **69,** 19–25.

Smith, K. L., V. de Vos, H. Bryden, L. B. Price, M. E. Hugh-Jones, and O. Keim, 2000: *Bacillus anthracis* diversity in Kruger National Park. J. Clin. Microbiol. **38,** 3780–3784.

Smith, D. L., B. Lucey, L. A. Waller, J. E. Childs, and L. A. Real, 2002: Predicting the spacial dynamics of rabies epidemics on heterogeneous landscapes. Proc. Natl Acad. Sci. U S A **99,** 3668–3672.

Tango, T., 2000: A test for spatial disease clustering adjusted for multiple testing. Stat. Med. **19,** 191–204.

Tango, T., and K. Takahashi, 2005: A flexibly shaped spatial scan statistic for detecting clusters. Int. J. Health Geogr. **4,** 11.

Ward, M. P., and T. E. Carpenter, 2000: Analysis of time–space clustering in veterinary epidemiology. Prev. Vet. Med. **43,** 225–237.

Ward, M. P., and A. M. Pérez, 2004: Association between soil type and paratuberculosis in cattle herds. Am. J. Vet. Res. **65,** 10–14.

Wilesmith, J. W., M. A. Stevenson, C. B. King, and R. S. Morris, 2003: Spatio-temporal epidemiology of foot-and-mouth disease in two counties of Great Britain in 2001. Prev. Vet. Med. **61,** 157–170.

Woolhouse, M., and A. Donaldson, 2001: Managing foot-and-mouth – the science of controlling disease outbreaks. Nature **410,** 515–516.